We can extend Central Limit Theorem to two populations as below

Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s with mean μ_1 and variance σ_1^2 and let $Y_1, Y_2, \ldots$ be i.i.d. r.v.'s with mean μ_2 and variance σ_2^2.

Let $\bar{X} = \frac{1}{m}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

Then the distribution of

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$$ converges to

$N(0,1)$ as $m \to \infty$ and $n \to \infty$.

<u>Examples</u>: 1. Let a random sample of size 54 from a discrete distribution with pmf $p(x) = \frac{1}{3}$, $x = 2, 4, 6$.

Find the prob that the sample mean will lie between 4.1 to 4.4.

We will apply the CLT here.

$$\mu = \frac{1}{3}(2+4+6) = 4,$$

$$E(X^2) = \frac{1}{3}\cdot(4+16+36) = \frac{56}{3}$$

$$\sigma^2 = \frac{56}{3} - 16 = \frac{8}{3}$$

By CLT $\quad \dfrac{\sqrt{54}\,(\bar{X}_{54} - 4)}{\sqrt{8/3}} \longrightarrow N(0,1)$

$$P(4.1 \le \bar{X}_{54} \le 4.4)$$

$$\approx P\left(\frac{\sqrt{54}\,(4.1-4)}{\sqrt{8/3}} \le Z \le \frac{\sqrt{54}\,(4.4-4)}{\sqrt{8/3}} \right)$$

$$= P(0.45 \le Z \le 1.8) = \Phi(1.8) - \Phi(0.45)$$

$$= 0.9641 - 0.6736 = 0.2905$$

2. The TV picture tubes of manufacturer A have a mean lifetime of 6.5 years and s.d. 0.9 years. Those from manufacturer B have a mean life of 6 years and a s.d. of 0.8 years. What is the prob. that a random sample of 36 tubes from A will have a mean life that is at least 1 year more

than the mean life of a random sample of 49 tubes from B?

$$P(\bar{X} - \bar{Y} > 1)$$

$\mu_1 = 6.5$, $\mu_2 = 6$, $\sigma_1 = 0.9$, $\sigma_2 = 0.8$

$m = 36$, $n = 49$ $\qquad\qquad$ $\mu_1 - \mu_2 = 0.5$

$$\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \cong 0.189,$$

$$\frac{\bar{X} - \bar{Y} - 0.5}{0.189} \rightarrow N(0,1)$$

$$P(\bar{X}-\bar{Y} > 1) \simeq P\left(Z > \frac{1-0.5}{0.189}\right)$$

$$= P(Z > 2.65) = 0.004$$

Other Sampling Distributions

Chi-square Distribution (χ^2)

A continuous r.v. W is said to have a Chi-square distribution with n degrees of freedom if it has pdf given by

$$f_W(w) = \frac{1}{2^{\frac{n}{2}}\,\Gamma\frac{n}{2}}\, e^{-\frac{w}{2}}\, w^{\frac{n}{2}-1}, \quad \begin{array}{l} w>0 \\ n>0 \end{array}$$

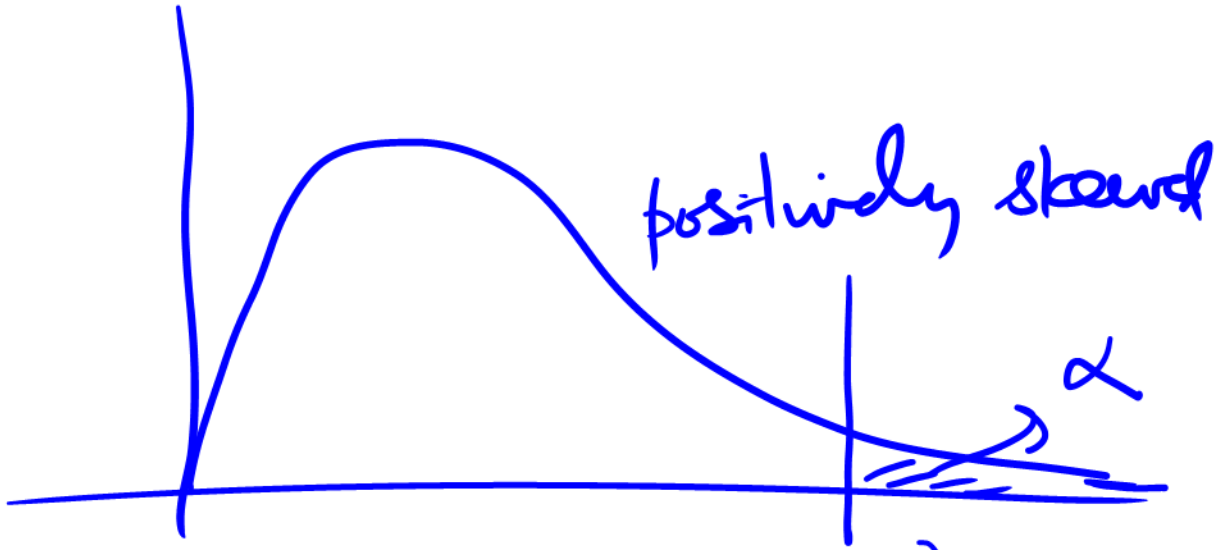
This is actually Gamma $\left(\frac{n}{2}, \frac{1}{2}\right)$ distn.

$E(W) = n, \quad V(W) = 2n$

$$\mu_k' = E(W^k) = n(n+2)\cdots\cdots(n+2(k-1))$$

$$M_W(t) = (1-2t)^{-\frac{n}{2}}, \quad t < \frac{1}{2}$$

$$\mu_3 = 8n > 0$$



positively skewed

$$\mu_4 = 12n(n+4)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{12}{n} > 0 \rightarrow \text{leptokurtic}$$

$\chi^2_{n,\alpha}$

$$P(W > \chi^2_{n,\alpha}) = \alpha$$

Additive Property of Chi-square dist^n

Let $W_1, W_2, \ldots, W_k$ be independently distributed with $W_i \sim \chi^2_{n_i}$. Then

$$U = \sum_{i=1}^{k} W_i \sim \chi^2_{\Sigma n_i}$$

Let $X \sim N(0,1)$ and $Y = X^2$

$x = \sqrt{y}$ $\qquad$ $x = -\sqrt{y}$

$\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$, $\qquad$ $\frac{dx}{dy} = -\frac{1}{2\sqrt{y}}$

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $\qquad -\infty < x < \infty$

The pdf of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2^{1/2}\Gamma\frac{1}{2}} e^{-\frac{y}{2}} y^{\frac{1}{2}-1}, \quad y > 0$$

which is pdf of χ^2_1 distn.

So if $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$ r.v's, then $W = \sum_{i=1}^{n} X_i^2 \sim \chi^2_n$

Next let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$M_{\bar{X}}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2n}}, \quad t \in \mathbb{R}$$

$$Y_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$$

$Y_1, \ldots, Y_n$ are i.i.d. $N(0,1)$

$$W = \sum_{i=1}^{n} Y_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi_n^2$$

$$M_W(t) = (1-2t)^{-n/2}, \qquad t < \frac{1}{2}$$

Let us define $U_i = X_i - \bar{X}$, $i = 1 \cdots n$

$$\underline{U} = (U_1, \ldots, U_n)$$

We first prove that $\bar{X}$ and $\underline{U}$ are independently distributed. We will use MGF approach, i.e., we will show that

$$M_{\bar{X}, \underline{U}}(s, \underline{t}) = M_{\bar{X}}(s)\, M_{\underline{U}}(\underline{t})$$

for all s, $\underline{t} = (t_1, \dots, t_n)$

Let $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$. We have

$$M_{\bar{X}}(s) = e^{\mu s + \frac{\sigma^2 s^2}{2n}} \quad \dots \quad (1)$$

Let us consider mgf of $\underline{U}$:

$$M_{\underline{U}}(\underline{t}) = E\left(e^{\sum_{i=1}^{n} t_i U_i}\right)$$

$$= E\left\{e^{\sum t_i (X_i - \bar{X})}\right\}$$

$$= E\left\{ e^{\sum t_i x_i - \bar{x} \sum t_i} \right\}$$

$$= E\left\{ e^{\sum t_i x_i - n \bar{t} \bar{x}} \right\}$$

$$= E\left\{ e^{\sum_{i=1}^{n} t_i x_i - \bar{t} \sum_{i=1}^{n} x_i} \right\}$$

$$= E\left\{ e^{\sum_{i=1}^{n} (t_i - \bar{t}) x_i} \right\}$$

$$= E\left\{\prod_{i=1}^{n} e^{(t_i-\bar{t})X_i}\right\}$$

$$= \prod_{i=1}^{n} E\left\{e^{(t_i-\bar{t})X_i}\right\}$$

$$= \prod_{i=1}^{n} M_{X_i}(t_i-\bar{t}) = \prod_{i=1}^{n}\left[e^{\mu(t_i-\bar{t})+\frac{1}{2}\sigma^2(t_i-\bar{t})^2}\right]$$

$$= e^{\frac{1}{2}\sigma^2 \sum_{i=1}^{n}(t_i - \bar{t})^2} \qquad \ldots (2)$$

Next, we consider

$$M_{\bar{X}, \underline{U}}(s, t) = E\left\{ e^{s\bar{X} + \sum_{i=1}^{n} t_i U_i} \right\}$$

$$= E\left\{ e^{\frac{s}{n}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}(t_i - \bar{t}) X_i} \right\}$$

$$= E\left\{ e^{\sum X_i \left(t_i - \bar{t} + \frac{s}{n}\right)} \right\}$$

$$= E\left[\prod_{i=1}^{n} e^{(t_i - \bar{t} + \frac{s}{n})X_i}\right]$$

$$= \prod_{i=1}^{n} E\left[e^{(t_i - \bar{t} + \frac{s}{n})X_i}\right]$$

$$= \prod_{i=1}^{n} M_{X_i}\left(t_i - \bar{t} + \frac{s}{n}\right)$$

$$= \prod_{i=1}^{n} \left\{ e^{\mu(t_i-\bar{t}) + \frac{s}{n} + \frac{1}{2}\sigma^2 \left(t_i-\bar{t}+\frac{s}{n}\right)^2} \right\}$$

$$= e^{\mu s + \frac{1}{2}\sigma^2 \sum (t_i-\bar{t})^2 + \frac{1}{2n}\sigma^2 s^2} \quad \cdots (3)$$

$$= M_{\bar{X}}(s)\, M_{\underline{U}}(\underline{t})$$

So $\bar{X}$ and $\underline{U} = (U_1, \ldots, U_n)$ are independently distributed.

Now consider $W = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left\{ (X_i - \bar{X}) + (\bar{X} - \mu) \right\}^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \quad \ldots (4)$$

$$= W_1 + W_2 \quad \text{(say)}$$

First we observe that W_1 and W_2 are independently distributed.

So $M_W(t) = M_{W_1}(t)\, M_{W_2}(t) \quad \cdots (5)$

$$W_2 = \frac{n(\bar{X}-\mu)^2}{\sigma^2}, \qquad W \sim \chi^2_n$$

$$\bar{X} \sim N(\mu, \sigma^2/n) \Rightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

$$\Rightarrow W_2 \sim \chi^2_1$$

So $\quad M_{W_1}(t) = M_W(t) \Big/ M_{W_2}(t)$

$$= \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-\left(\frac{n-1}{2}\right)}, \quad t < \frac{1}{2}$$

This mgf of χ^2_{n-1} dist^n

So we have proved that

$$W_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Define $S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \overline{X})^2$ as the

sample variance of $X_1, \ldots, X_n$.

Then we have proved that

$$\frac{(n-1) S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Student's t-distribution

Let X and Y be independently distributed random variables.

Let $X \sim N(0,1)$ and $Y \sim \chi^2_n$.

Then $T = \dfrac{X}{\sqrt{Y/n}}$ is said to have

a student's t-dist^n on n d.f.

$$T \sim t_n.$$

The joint pdf of X and Y is

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}, \quad x \in \mathbb{R},\ y > 0$$

Let $T = \frac{\sqrt{n}\,X}{\sqrt{Y}}$ and $U = Y$

The inverse transformation is

$$x = \sqrt{\frac{u}{n}}\, t, \qquad y = u$$

The Jacobian of transformation is

$$J = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial u} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{u}{n}} & \frac{t}{2\sqrt{nu}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}}$$

The joint pdf of T and U is

$$f_{T,U}(t,u) = \frac{1}{2^{\frac{n+1}{2}}\sqrt{\pi n}\,\Gamma\frac{n}{2}}\, e^{-\frac{u}{2}\left(1+\frac{t^2}{n}\right)}\, u^{\frac{n+1}{2}-1}, \quad t \in \mathbb{R},\ u > 0$$

The marginal pdf of T is

$$f_T(t) = \frac{\Gamma\frac{n+1}{2}}{\sqrt{n}\,\Gamma\frac{n}{2}\,\Gamma\frac{1}{2}} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

$$= \frac{1}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

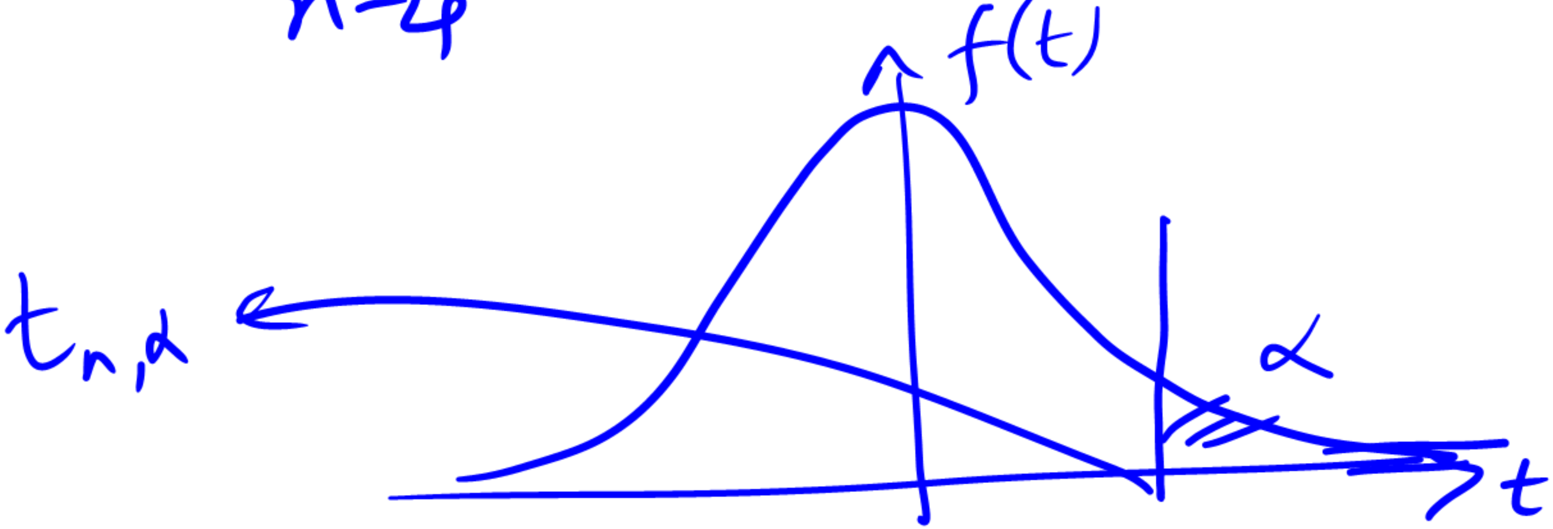
The density is symmetric about 0.
So odd ordered moments will vanish (which exist)

Even ordered moments exist of order $(<n)$

$$E(T) = 0, \quad V(T) = E(T^2) = \frac{n}{n-2}, \quad n>2$$

$$\mu_4 = E(T^4) = \frac{3n^2}{(n-2)(n-4)}, \quad n>4$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6}{n-4} > 0 \quad \text{leptokurtic}$$



$$P(T > t_{n,\alpha}) = \alpha$$

As $n \to \infty$, the pdf T converges to $\phi(t)$ (i.e. pdf of standard normal r.v.).

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$

Then $\sqrt{n}\left(\frac{\overline{X} - \mu}{\sigma}\right) \sim N(0,1)$

and $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

and these two are independently distributed

So $$\frac{\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$$

Snedcor's F-distribution

Let W_1 and W_2 be independent r.v.'s

and $W_1 \sim \chi^2_m$, $W_2 \sim \chi^2_n$.

Then $U = \dfrac{(W_1/m)}{(W_2/n)}$ is said to have

an F-distⁿ on (m, n) d.f.

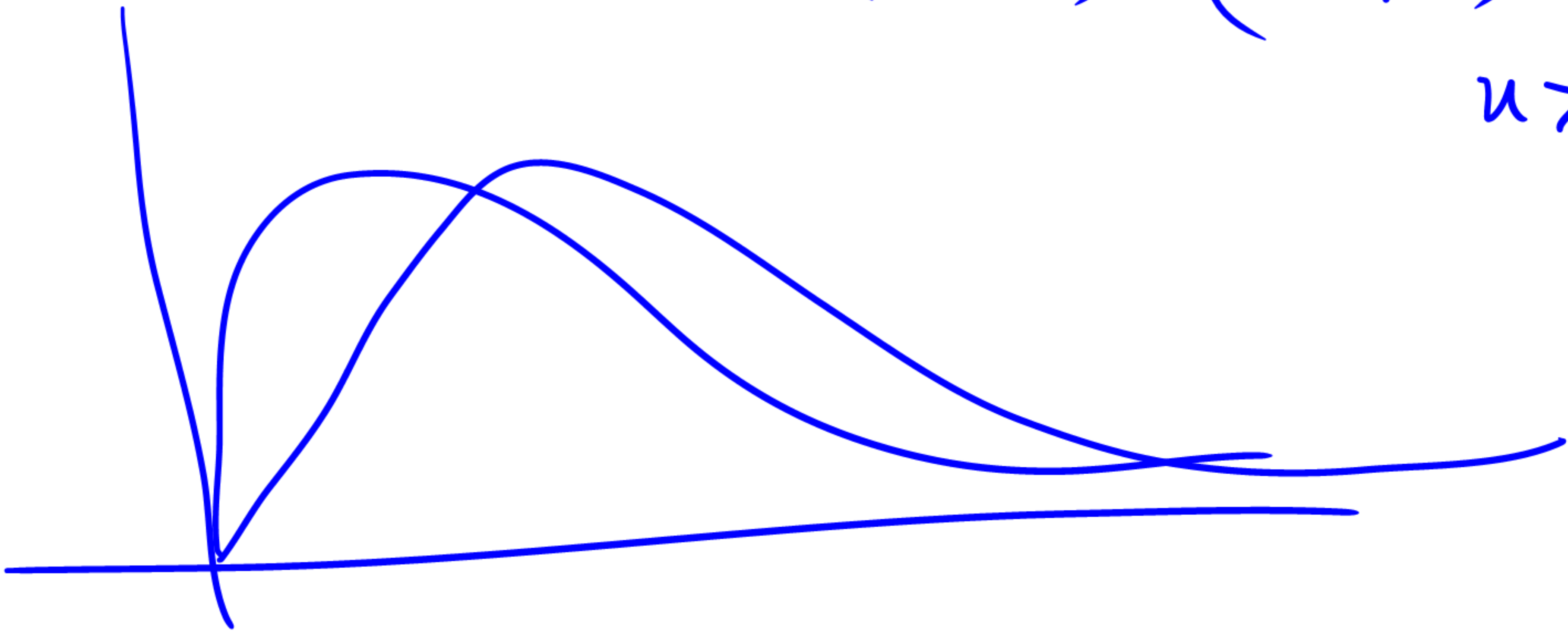
$$U \sim F_{m,n}$$

⊛ Derive the pdf of U.

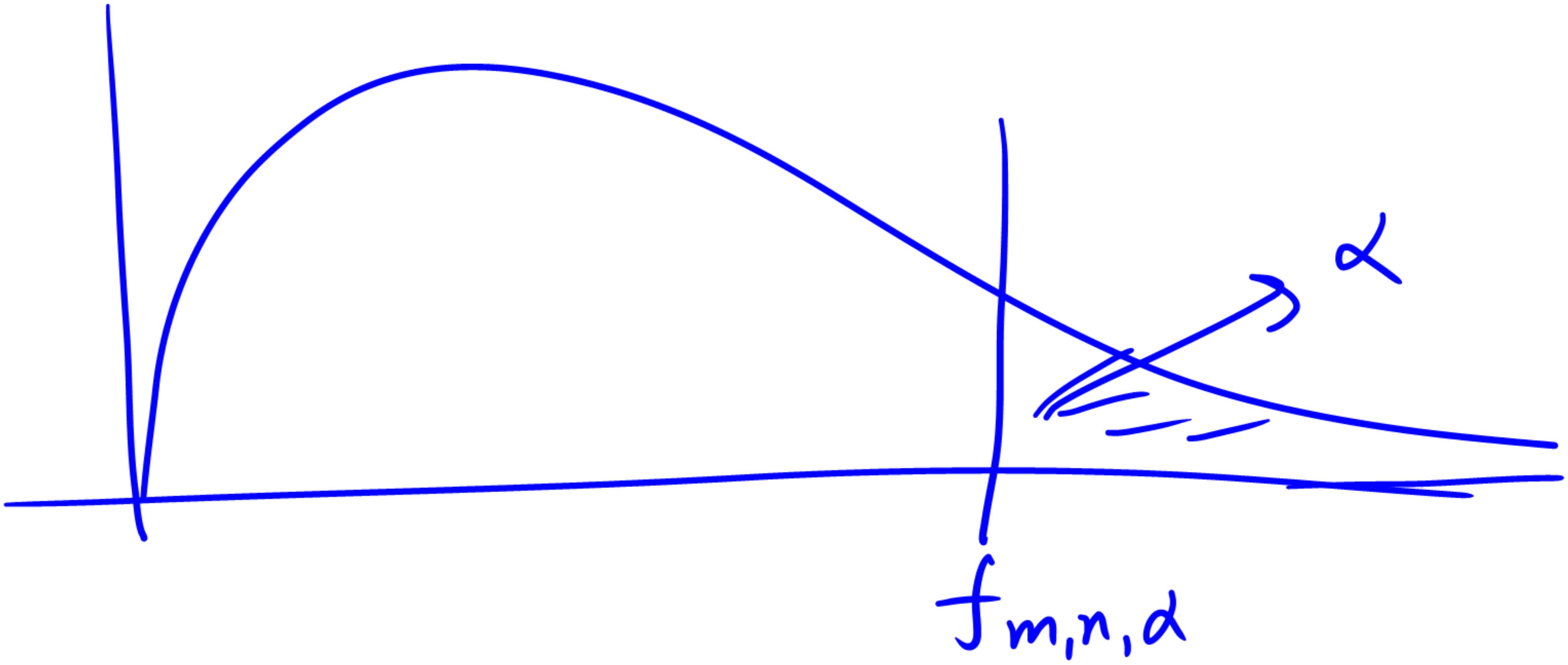
The pdf of U is

$$f_U(u) = \left(\frac{m}{n}\right)^{m/2} \cdot \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{u^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}u\right)^{\frac{m+n}{2}}},$$

$u > 0$



$$E(U) = \frac{n}{n-2}, \quad n > 2, \qquad V(U) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n > 4.$$



$$P(U > f_{m,n,\alpha}) = \alpha$$

If $U \sim F_{m,n}$ then $\frac{1}{U} \sim F_{n,m}$

So $f_{1-\alpha, n, m} = \frac{1}{f_{\alpha, m, n}}$.

Let $X_1, \ldots, X_m$ be a random sample from $N(\mu_1, \sigma_1^2)$ and let $Y_1, \ldots, Y_n$ be a random sample from $N(\mu_2, \sigma_2^2)$.

$$S_1^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} (X_i - \bar{X})^2, \qquad S_2^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (Y_j - \bar{Y})^2$$

$$W_1 = \frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}, \qquad W_1 \text{ \& } W_2 \text{ are indept.}$$

$$W_2 = \frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

Then $$U = \frac{W_1/(m-1)}{W_2/(n-1)} = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2} \sim F_{m-1,n-1}$$